



**PHILIPS**

*Let's make things better.*

# MPEG-7: Visual Part

*Sylvie Jeannin*

*[sylvie.jeannin@philips.com](mailto:sylvie.jeannin@philips.com)*



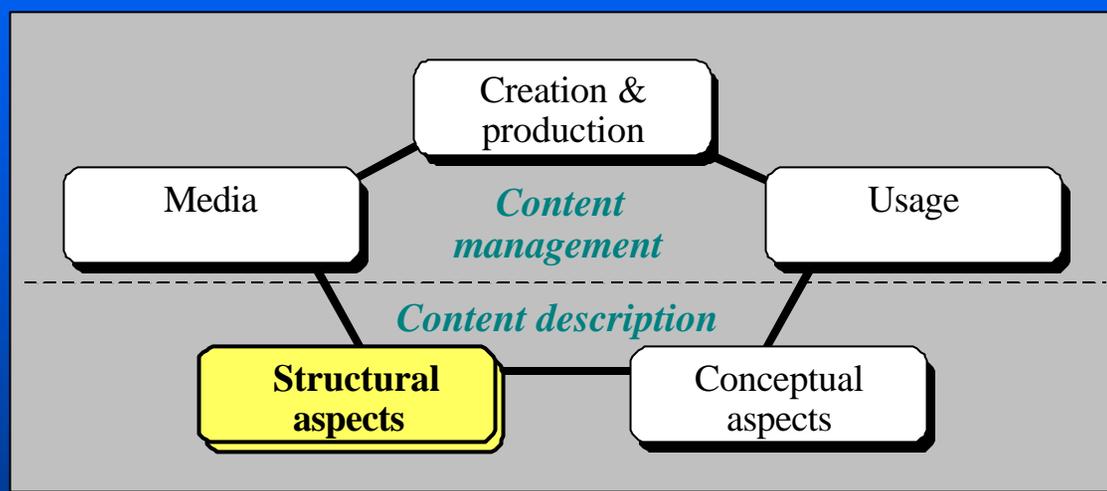
**PHILIPS**

# Outline

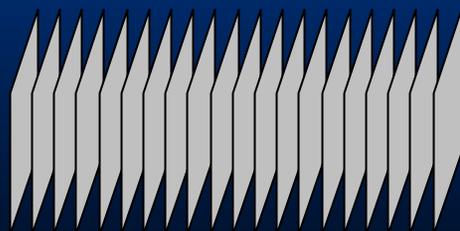
- **Introduction on MPEG-7 Visual Part**  
context and goals, structure...
- **Overview of the Visual Part**  
feature by feature, Descriptor by Descriptor
- **Conclusion and References**

# Introduction on MPEG-7 Visual Part

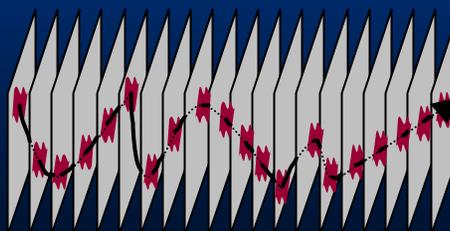
## *MPEG-7 Visual part: Context*



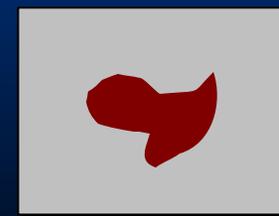
### Video segments



### Moving regions



### Still regions



# Introduction on MPEG-7 Visual Part

## *MPEG-7 Visual part: Evaluation, Selection*

**Core Experiments:** per feature (context) / functionality

- technical proposal(s)
- 2 independent parties
- test set
- functionality(s) (broad / generic, demonstrative)
- evaluation criteria defined depending on the CE
  - compactness
  - automatic / simple extraction
  - simple use
  - expressiveness, efficiency, ...

# Introduction on MPEG-7 Visual Part

## *Structure of a Descriptor*

### Normative Parts

#### DDL specification

```
<complexType name="MyType">  
<complexContent>  
<extension  
  base="mpeg7:VisualDType">  
<sequence>  
<element name="FirstElement"...
```

#### Binary format

	Bits
FirstElement	1
...	...

#### Fields semantics

##### FirstElement

This field is a  
boolean  
representing ...

### Non-Normative Parts

- Recommended extraction method(s)
- Recommended usage method(s) for selected functionalities
- Conditions of use

# Introduction on MPEG-7 Visual Part

## *MPEG-7 Visual part: Common Properties*

### **One tool for one given context / functionality**

- broad and generic: not application centric
- allowing interoperability
- with extraction method: automatic, simple
- with normative representation: compact
- with recommended usage

# Introduction on MPEG-7 Visual Part

## *MPEG-7 Visual part components*

***MPEG-7 Visual part contains 25 Ds/DSs***

**Basic Elements (2)**

**Color (7)**

**Texture (3)**

**Shape (3)**

**Motion (4)**

**Localization (2)**

**Face (1)**

**Containers (3)**

# Overview of the Visual Part

## *Basic Elements*

### • Coordinates:

### *Spatial2DCoordinates*

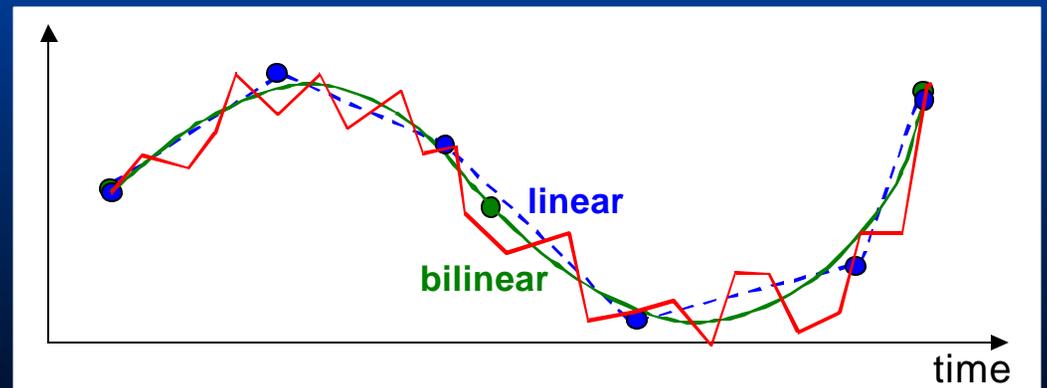
- Units: meters, pixels, pixels normalized by image size
- Spatial Reference: fixed, varying (to follow global motion)

➡ *independence of descriptions from display format*

### • Temporal Interpolation:

### *TemporalInterpolation*

- (n-Dimension + time) values
- optionally parameters define bilinear interpolation



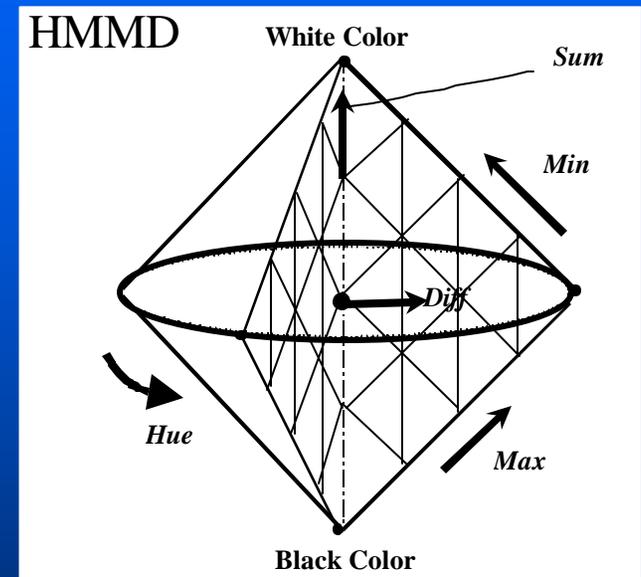
# Overview of the Visual Part

## Color (1)

### • Color Space:

- R, G, B
- Y, Cr, Cb
- H, S, V
- Monochrome
- Linear transformation of R, G, B
- HMMD

### ColorSpace



### • Quantization:

### ColorQuantization

uniform quantization scheme,  
to be applied on any of the color spaces defined above

# Overview of the Visual Part

## *Color (2)*

- **Dominant Color(s):** *DominantColor*

1-8 dominant colors in image / region

- color space, quantization, dominant color(s) value(s)
- variance of color value, percentage of pixels of this color, spatial coherency of color repartition

- **Color Content (histogram):** *ScalableColor*

Color histogram transformed by Haar transform

- scalable in number of coefficients kept for representation
- scalable in number of bits per coefficients
- lower end: 60 bits, very fast matching

# Overview of the Visual Part

## *Color (3)*

• **Color content + coherence of colors repartition:** *ColorStructure*

Histogram of structuring elements that contain a particular color.

➡ *Enhanced retrieval (in conjunction with HMMD)*

• **Color content + its layout:** *ColorLayout*

Based on the DCT coefficients. (size: about 160 bits)

➡ *Layout sensitive retrieval, sketch-to-image matching*

• **Color content of Group of Pictures / Frames:** *GoFGoPColor*

Aggregation of color histograms (average, median, or intersection)

➡ *Clustering of data for browsing / retrieval*

# Overview of the Visual Part

## *Texture (1)*

- **Characterization of homogeneous textures:**

- low-level: *HomogeneousTexture*      ⇒ *retrieval*
- high level: *TextureBrowsing*      ⇒ *browsing*

- **Characterization of structures in generic images:**

- edges content and layout: *EdgeHistogram*

# Overview of the Visual Part

## Texture (2)

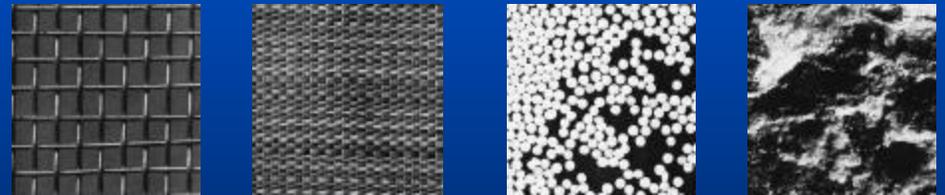
Low-level: *HomogeneousTexture*

High-level: *TextureBrowsing*

In the frequency domain:

- decomposition onto 30 channels (5 scales, 6 angles) using Gabor filters
- energy and energy deviation

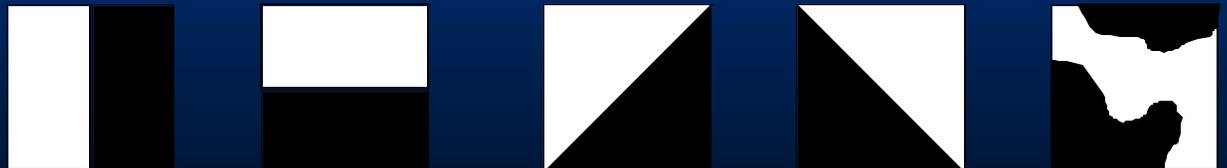
- regularity (1 to 4)
- main direction(s)
- coarseness (1 to 4)



*EdgeHistogram*

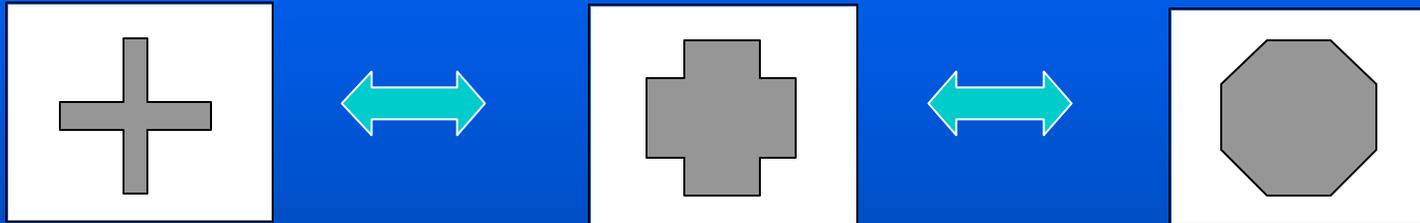
- Edge types histograms on 16 sub-images

- Fixed size: 240 bits



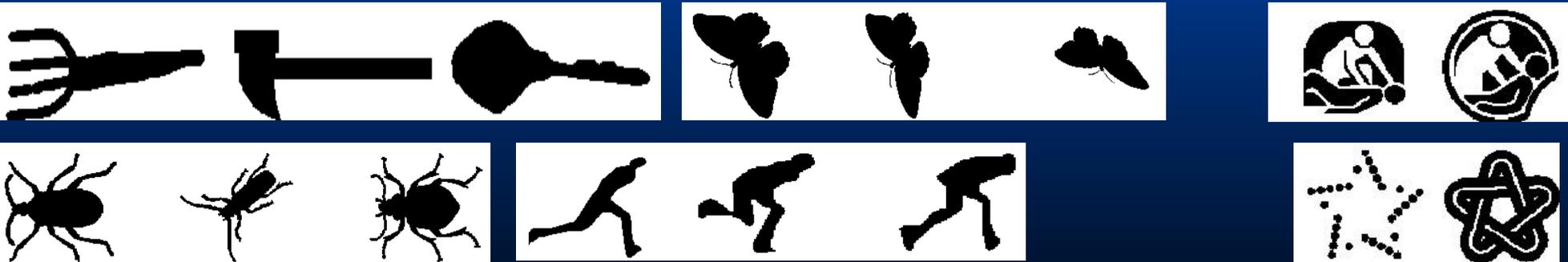
# Overview of the Visual Part *Shape (1)*

- **2D:** *ContourShape and RegionShape*



- **3D:** *Shape3D*

- **Some 2D similarity-based retrieval results examples**



# Overview of the Visual Part *Shape (2)*

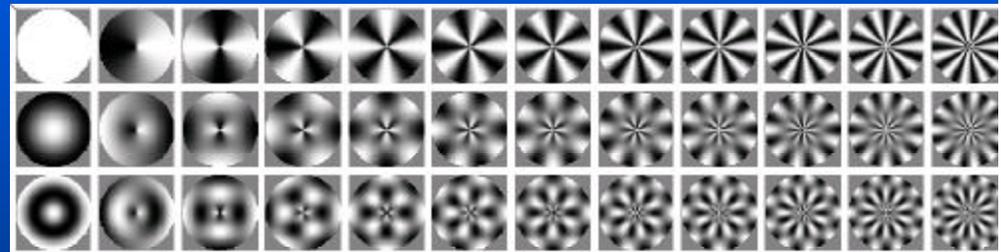
Contour Based: *ContourShape*

- Curvature Scale Space:  
curvature points importance  
and relative positions

- Variable size: < 15 Bytes

Region Based: *RegionShape*

- Art moments



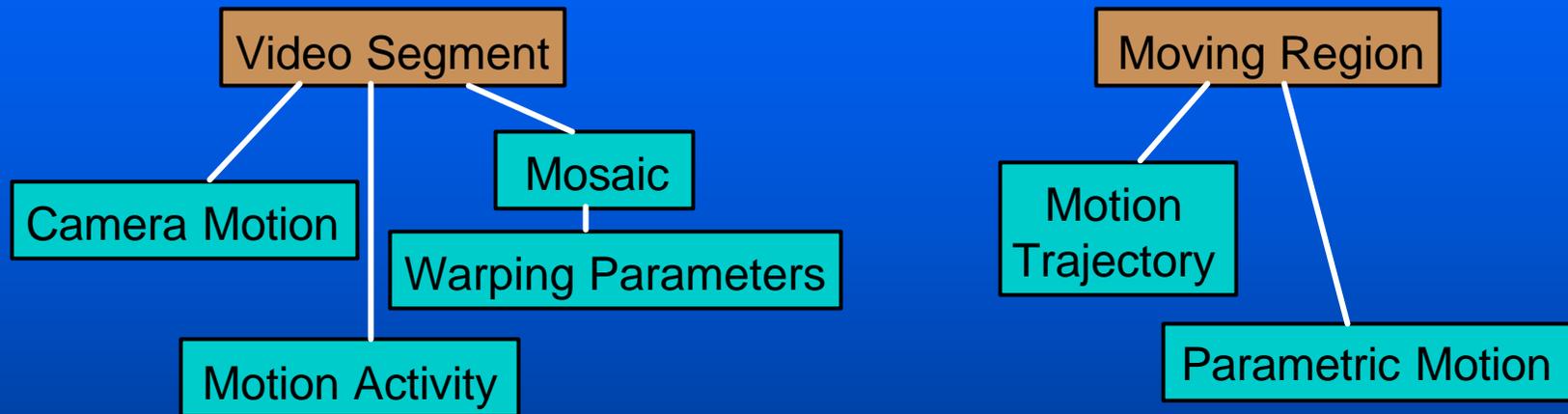
- Fixed size: 17.5 Bytes

*Shape3D*

- based on 3D meshes
- histogram of 3D shape indexes, which represent local curvature properties of the 3D surface

# Overview of the Visual Part

## *Motion (1)*



### *MotionActivity*

browsing, repurposing

### *CameraMotion*

browsing, high level queries

### *MotionTrajectory*

retrieval, high level queries

### *ParametricMotion*

mosaic, retrieval

# Overview of the Visual Part

## Motion (2)

- Motion Activity:**

### MotionActivity

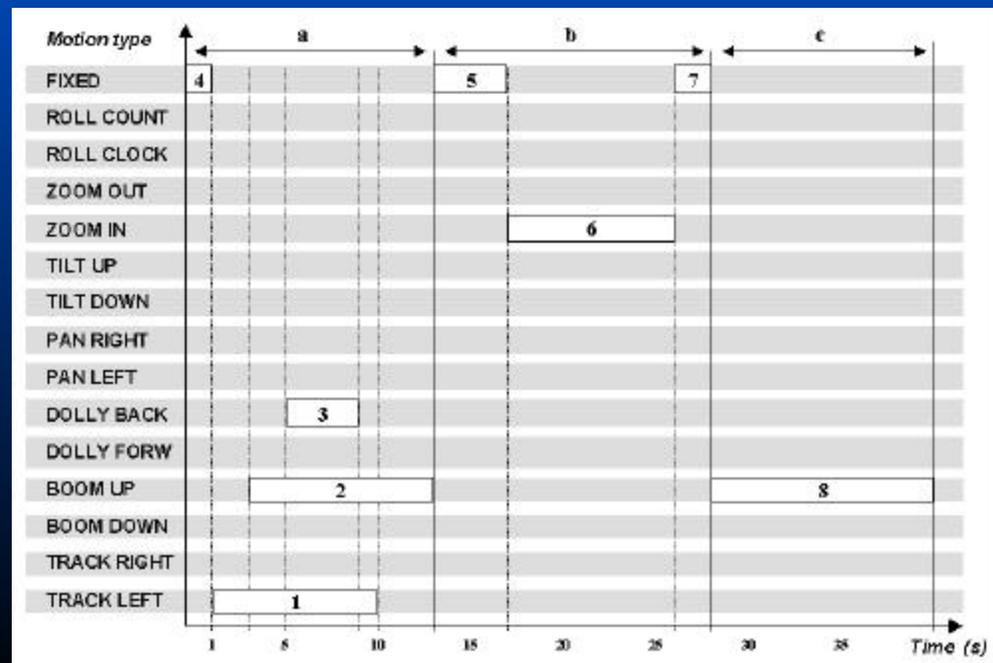
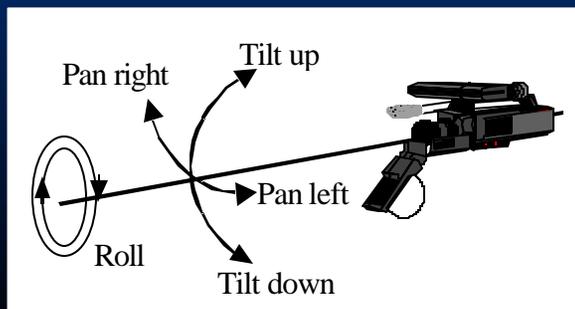
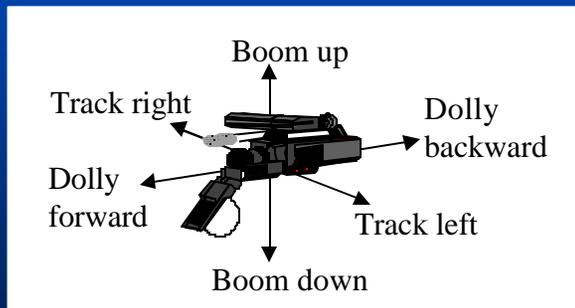
Intensity of motion (1 to 5)

main direction  
spatial localization in image

temporal repartition in Segment  
type of active regions

- Camera Motion:**

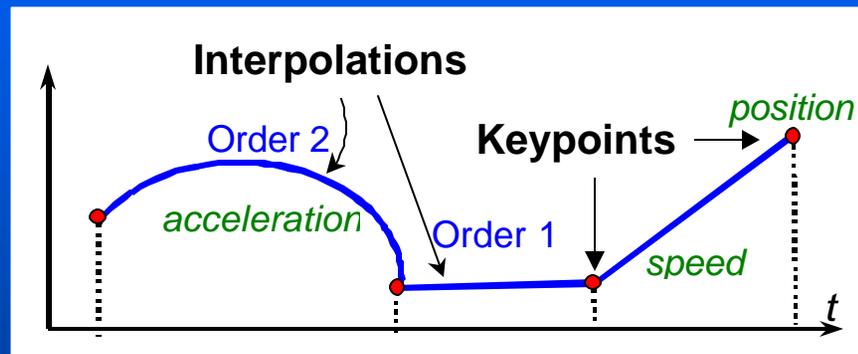
### CameraMotion



# Overview of the Visual Part *Motion (3)*

## • Motion Trajectory:

## *MotionTrajectory*



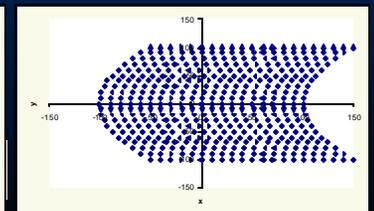
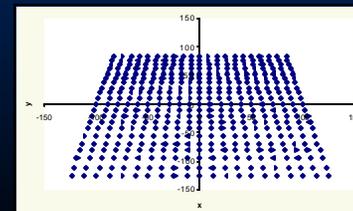
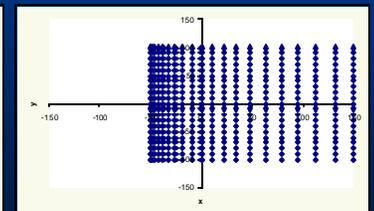
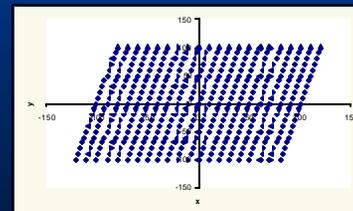
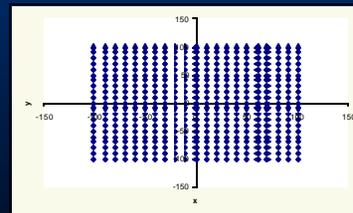
## Queries:

- similarity
- high level

## • Parametric Motion:

## *ParametricMotion*

- translational
- rotation/scaling
- affine
- planar perspective
- parabolic



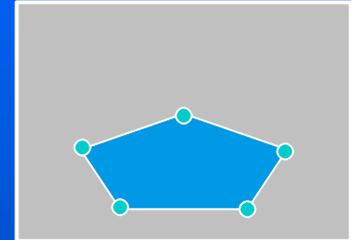
# Overview of the Visual Part

## Localization

- **Spatial Localization:**

*RegionLocator*

Approximation of region by box or polygon  
Position of vertices in image

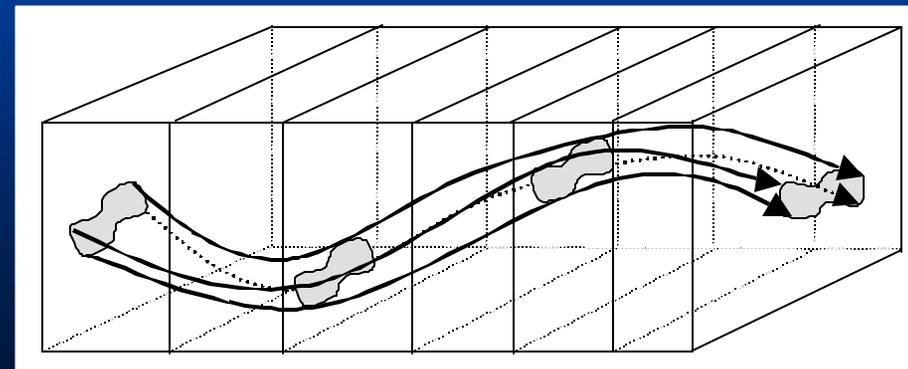


- **Spatio-Temporal Localization:**

*SpatioTemporalLocator*

Reference region, approximated by box, ellipse or polygon.

- trajectories of representative points
- temporal interpolation of parametric motion parameters



 *hyperlinking*

# Overview of the Visual Part *Face*

- **Face Characterization:**

- *FaceRecognition*

- size: 238 bits

- based on eigenfaces (vector of 2576 values, extracted from normalized faces)

- 49 basis vectors which span the space of possible face vectors

- projection of the face vector on the 49 eigenfaces

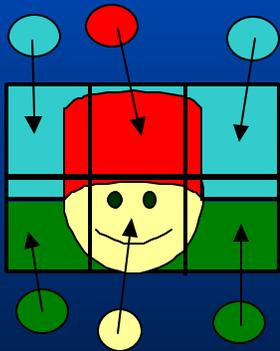
# Overview of the Visual Part *Containers*

spatial

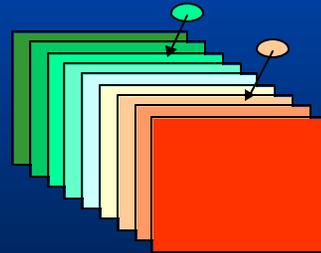
temporal

2D-3D

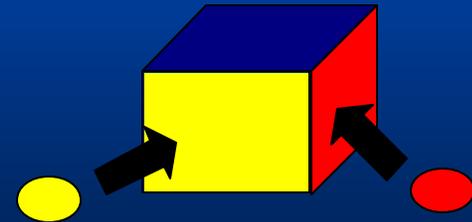
*GridLayout*



*TimeSeries*



*MultipleViews*



# Conclusion

## **MPEG-7 Visual parts contains 25 Ds/DSs**

- characterizing images, video segments, regions, 3D objects, faces
- in terms of color, texture, shape, motion, localization, face features
- one by one, or gathered spatially (grid, views) or temporally
- automatically extractable from visual content
- of compact size
- allowing similarity-based retrieval, browsing, high-level queries ...

# References

**For more details:**

**<http://www.csel.it/mpeg/>**

- **MPEG-7 Overview Document**
- **MPEG-7 Visual FCD and XM**